SCALING

Prof. Dan A. Simovici

UMB

Scaling is a process that allows us to represent high-dimensional spaces in low-dimensional Euclidean spaces, by conserving the distances between the original points as much as possible.
Scaling is important for visualizing the result of data explorations.

Let $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ be a sequence of m vectors in $\mathbb{R}^n$. The corresponding matrix is

$$X = (\mathbf{x}_1, \ldots, \mathbf{x}_m) \in \mathbb{R}^{n \times m}.$$

Note that X is the transpose of the sample data matrix previously considered (which had $\mathbf{x}_1', \ldots, \mathbf{x}_m'$ as its rows).
Given the matrix of Euclidean distances $D = (d_{ij}^2) \in \mathbb{R}^{m \times m}$, where

$$d_{ij}^2 = \| \mathbf{x}_i - \mathbf{x}_j \|_2^2 = (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j)$$

for $1 \leq i, j \leq m$, we need to retrieve the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$.
This problem does not have a unique solution because the matrix D is the same for $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ and for $(\mathbf{x}_1 + \mathbf{c}, \ldots, \mathbf{x}_m + \mathbf{c})$ for every $\mathbf{c} \in \mathbb{R}^n$.

Let $G \in \mathbb{R}^{m\times m}$ be the Gram matrix of X, $G = G_X = X'X$. Since $g_{pq} = \mathbf{x}'_p\mathbf{x}_q$ we have

$$d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j) = g_{ii} + g_{jj} - 2g_{ij} \tag{1}$$

for $1 \leq i, j \leq m$.
Suppose now that $F \in \mathbb{R}^{m\times m}$ is the Gram matrix of another sequence of vectors $(\mathbf{y}_1, \ldots, \mathbf{y}_m)$ having the same distaces, that is, $F = G_Y = Y'Y$, where $Y = (\mathbf{y}_1, \ldots, \mathbf{y}_m) \in \mathbb{R}^{n\times m}$ such that $d_{ij}^2 = f_{ii} + f_{jj} - 2f_{ij}$. Then, $g_{ii} + g_{jj} - 2g_{ij} = f_{ii} + f_{jj} - 2f_{ij}$ for $1 \leq i, j \leq m$.

Let $W = G - F$. Then, W is a symmetric matrix and $w_{ii} + w_{jj} - 2w_{ij} = 0$, so $w_{ij} = \frac{1}{2}(w_{ii} + w_{jj})$. Let

$$\mathbf{w} = \frac{1}{2} \begin{pmatrix} w_{11} \\ \vdots \\ w_{mm} \end{pmatrix}$$

and note that the matrix W can now be written as $W = \mathbf{w}\mathbf{1}'_m + \mathbf{1}_m\mathbf{w}'$, which proves that W is a special, rank-2 matrix.

Consequently,

$$G = F + \mathbf{w}\mathbf{1}'_m + \mathbf{1}_m\mathbf{w}'. \tag{2}$$

Thus, the set of vectors that correspond to a distance matrix is not unique, and the Gram matrices of any two such sequences differ by a symmetric matrix of rank 2.

Construction of the Data Set X Starting from Distances

It is possible to construct X starting from D if we assume that the centroid of the vectors of X is $\mathbf{0}_n$, that is, if $\sum_{i=1}^{m} \mathbf{x}_i = \mathbf{0}_n$.
Let $A \in \mathbb{R}^{m \times m}$ be the matrix defined by $A = -\frac{1}{2}D$. Elementwise, this means that $a_{ij} = -\frac{1}{2}d_{ij}^2$ for $1 \leq i, j \leq m$. Consider the averages defined by:

$$
\begin{aligned}
a_{i\cdot} &= \frac{1}{m}\sum_{j=1}^{m} a_{ij}, \\
a_{\cdot j} &= \frac{1}{m}\sum_{i=1}^{m} a_{ij}, \\
a_{\cdot\cdot} &= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij}.
\end{aligned}
$$

The components of the Gram matrix $G \in \mathbb{C}^{m \times m}$, $g_{ij} = \mathbf{x}_i' \mathbf{x}_j$ for $1 \leq i, j \leq m$, can be expressed using these averages, assuming that the set of columns of X is centered in $\mathbf{0}_n$.

Theorem

Let $X = (\mathbf{x}_1, \ldots, \mathbf{x}_m) \in \mathbb{R}^{n \times m}$ *be a matrix such that* $\sum_{i=1}^{m} \mathbf{x}_i = \mathbf{0}_n$ *and let* A *be the matrix defined by* $a_{ij} = -\frac{1}{2} \parallel \mathbf{x}_i - \mathbf{x}_j \parallel_2^2$ *for* $1 \leq i, j \leq m$. *The components of the Gram matrix* G, $g_{ij} = \mathbf{x}_i' \mathbf{x}_j$ *are given by*

$$g_{ij} = a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot}$$

for $1 \leq i, j \leq m$.

Proof

By Equality (1) we have

$$-2a_{ij} = g_{ii} + g_{jj} - 2g_{ij}$$

for $1 \leq i, j \leq m$. Note that

$$\sum_{i=1}^{m} g_{ij} = \sum_{j=1}^{m} g_{ij} = 0.$$

The averages introduced earlier can be written as

$$\begin{aligned} -2a_{\cdot j} &= \frac{1}{m}\sum_{i=1}^{m} d_{ij}^2 = \frac{1}{m}\sum_{i=1}^{m}(g_{ii} + g_{jj} - 2g_{ij}) \\ &= \frac{1}{m}\sum_{i=1}^{m} g_{ii} + g_{jj} - 2\left(\frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i'\right)\mathbf{x}_j \\ &= g_{jj} + \frac{1}{m}\sum_{i=1}^{m} g_{ii}, \end{aligned}$$

Equality (1) yields

$$
\begin{aligned}
g_{ij} &= -\frac{1}{2}\left(d_{ij}^2 - g_{ii} - g_{jj}\right) \\
&= -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{m}\sum_{j=1}^{m} d_{ij}^2 + \frac{1}{m}\sum_{j=1}^{m} \mathbf{x}_j\mathbf{x}_j' - \frac{1}{m}\sum_{j=1}^{m} d_{ij}^2 + \frac{1}{m}\sum_{i=1}^{m} \mathbf{x}_i\mathbf{x}_i'\right) \\
&= -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{m}\sum_{j=1}^{m} d_{ij}^2 - \frac{1}{m}\sum_{j=1}^{m} d_{ij}^2 + \frac{1}{r^2}\sum_{i=1}^{m}\sum_{j=1}^{m} d_{ij}^2\right) \\
&= a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot},
\end{aligned}
$$

which completes the proof.

Corollary

The Gram matrix $G = X'X$ *of the sequence of* $\mathbb{R}^n$ *vectors* $X = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$ *can be obtained from the matrix* $A = \left(-\frac{1}{2} \parallel \mathbf{x}_i - \mathbf{x}_j \parallel_2^2\right)$ *as* $G = H_m A H_m$, *where* H_m *is the centering matrix* $H_m = I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}'_m$.

Proof

The matrix $H_m A H_m$ can be written as

$$\begin{aligned} H_m A H_m &= \left(I_m - \frac{1}{m} \mathbf{1}_m \mathbf{1}'_m \right) A \left(I_m - \frac{1}{m} \mathbf{1}_m \mathbf{1}'_m \right) \\ &= \left(I_m - \frac{1}{m} \mathbf{1}_m \mathbf{1}'_m \right) \left(A - \frac{1}{m} A \mathbf{1}_m \mathbf{1}'_m \right) \\ &= A - \mathbf{1}_m \left(\frac{1}{m} \mathbf{1}'_m A \right) - \left(\frac{1}{m} A \mathbf{1}_m \right) \mathbf{1}' + \frac{1}{m^2} \mathbf{1}_m \left(\mathbf{1}'_m A \mathbf{1}_m \right) \mathbf{1}'_m. \end{aligned}$$

The terms of the above sum correspond to a_{ij}, $a_{\cdot j}$, $a_{i\cdot}$ and $a_{\cdot\cdot}$, respectively. The desired conclusion follows.

- We have $rank(G) = rank(X)$ because $G = X'X$, so $rank(G) = n$.
- Since G is symmetric, positive semi-definite and of rank n, it follows that G has n non-negative eigenvalues and $m - n$ zero eigenvalues.

We have $G = U'DU$, where U is an orthogonal matrix, $U = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_m)$, $D = (\lambda_1, \ldots, \lambda_n, 0, \ldots, 0)$ and $\lambda_1 \geq \cdots \geq \lambda_n > 0$.
Eliminating the $m - n$ elements of the diagonal of G which are 0 we can write $G = V'\text{diag}(\lambda_1, \ldots, \lambda_n)V$, where $V \in \mathbb{R}^{n \times m}$.
By defining X as

$$X = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})V \in \mathbb{R}^{n \times m}$$

we have $G = X'X$ and the m columns of X yield the desired vectors in $\mathbb{R}^n$.

Example

For the matrix

$$X = \begin{pmatrix} 5 & 5 & 8 & 6 \\ 8 & 4 & 6 & 6 \\ 0 & 0 & 0 & 6 \end{pmatrix}$$

representing four points $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ and $\mathbf{x}_4$ in $\mathbb{R}^3$ we have

$$G = X'X = \begin{pmatrix} 89 & 57 & 88 & 78 \\ 57 & 41 & 64 & 54 \\ 88 & 64 & 100 & 84 \\ 78 & 54 & 84 & 108 \end{pmatrix}$$

G can be factored as $G = UDU'$ using the MATLAB function, `[U,D] = eig(G)`, where U is an orthogonal matrix of eigenvectors. Because G is symmetric we have

$$G = \begin{pmatrix} -0.0506 & 0.7845 & 0.3373 & 0.5178 \\ 0.8608 & -0.2795 & 0.2294 & 0.3581 \\ -0.5064 & -0.5535 & 0.3563 & 0.5570 \\ 0.0000 & 0.0039 & -0.8406 & 0.5416 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 6.9950 & 0 & 0 \\ 0 & 0 & 26.3535 & 0 \\ 0 & 0 & 0 & 304.6515 \end{pmatrix} \cdot$$
$$\begin{pmatrix} -0.0506 & 0.8608 & -0.5064 & 0.0000 \\ 0.7845 & -0.2795 & -0.5535 & 0.0039 \\ 0.3373 & 0.2294 & 0.3563 & -0.8406 \\ 0.5178 & 0.3581 & 0.5570 & 0.5416 \end{pmatrix}$$

As expected, D has only three non-zero eigenvalues.

Let $E \in \mathbb{R}^{3\times 3}$ be the matrix

$$E = \begin{pmatrix} \sqrt{\lambda_2} & 0 & 0 \\ 0 & \sqrt{\lambda_3} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{pmatrix} = \begin{pmatrix} 2.6448 & 0 & 0 \\ 0 & 5.1336 & 0 \\ 0 & 0 & 17.4543 \end{pmatrix},$$

where we retained only the non-zero eigenvalues. Then, a set of points that has the same Gram matrix, and, therefore, the same inter-distances is given by the columns of the matrix

$$X = E * U(:, 2:4)' = \begin{pmatrix} 2.0749 & -0.7391 & -1.4640 & 0.0103 \\ 1.7316 & 1.1779 & 1.8292 & -4.3153 \\ 9.0386 & 6.2503 & 9.7217 & 9.4540 \end{pmatrix}$$

It is straingforward to verify that $X'X$ yields again G.

If we wish to sacrifice precision to reduce dimensionality we may drop not just 0 but the least two eigenvalues, by taking

$$E = \begin{pmatrix} 5.1336 & 0 \\ 0 & 17.4543 \end{pmatrix},$$

and

$$X = E * U(:, 3:4)' = \begin{pmatrix} 1.7316 & 1.1779 & 1.8292 & -4.3153 \\ 9.0386 & 6.2503 & 9.7217 & 9.4540 \end{pmatrix}.$$

The Gram matrix $G = X'X$ is

$$\begin{pmatrix} 84.6950 & 58.5338 & 91.0380 & 77.9788 \\ 58.5338 & 40.4539 & 62.9182 & 54.0077 \\ 91.0380 & 62.9182 & 97.8571 & 84.0153 \\ 77.9788 & 54.0077 & 84.0153 & 108.0005 \end{pmatrix}.$$

A further sacrifice of precision retains only the largest eigenvalue. In this case,

$$X = 17.4543 * U(:,4)' = (9.0378 \;\; 6.2504 \;\; 9.7220 \;\; 9.4532).$$

The matrix $G = X'X$ is

$$\begin{pmatrix} 81.6825 & 56.4900 & 87.8663 & 85.4369 \\ 56.4900 & 39.0673 & 60.7665 & 59.0864 \\ 87.8663 & 60.7665 & 94.5182 & 91.9049 \\ 85.4369 & 59.0864 & 91.9049 & 89.3639 \end{pmatrix}$$

A more general problem begins with a matrix of dissimilarities $\Delta = (\delta_{ij}) \in \mathbb{R}^{m \times m}$ and seeks to determine whether there exists a sequence of vectors $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ in $\mathbb{R}^n$ such that $d(\mathbf{x}_i, \mathbf{x}_j) = \delta_{ij}$ for $1 \leq i, j \leq m$.

Lemma

Let $A, G \in \mathbb{R}^{m\times m}$ be two matrices such that $G = H_m A H_m$, where H_m is the centering matrix $H_m = I_m - \frac{1}{n}\mathbf{1}_m\mathbf{1}'_m$. Then,

$$g_{ii} + g_{jj} - 2g_{ij} = a_{ii} + a_{jj} - 2a_{ij}$$

for $1 \leq i, j \leq m$.

Proof

We saw that if $G = H_m A H_m$, then $g_{ij} = a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot}$. Therefore, we have

$$\begin{aligned} g_{ii} &= a_{ii} - 2a_{i\cdot} + a_{\cdot\cdot}, \\ g_{jj} &= a_{jj} - 2a_{\cdot j} + a_{\cdot\cdot}. \end{aligned}$$

This allows us to write

$$\begin{aligned} g_{ii} + g_{jj} - 2g_{ij} &= a_{ii} - 2a_{i\cdot} + a_{\cdot\cdot} + a_{jj} - 2a_{\cdot j} + a_{\cdot\cdot} \\ &\quad -2(a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot}) \\ &= a_{ii} + a_{jj} - 2a_{ij}. \end{aligned}$$

Theorem

Let $\Delta \in \mathbb{R}^{m\times m}$ be a matrix of dissimilarities, $A \in \mathbb{R}^{m\times m}$ be the matrix defined by $a_{ij} = -\frac{1}{2}\delta_{ij}^2$ for $1 \leq i, j \leq m$, and let G be the centered matrix $G = H_m A H_m$. If G is a positive semi-definite matrix and rank$(G) = n$, then there exists a sequence $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ of vectors in $\mathbb{R}^n$ such that $d(\mathbf{x}_i, \mathbf{x}_j) = \delta_{ij}$ for $1 \leq i, j \leq m$.

Since G is a symmetric, positive semi-definite matrix having rank n, it is possible to write

$$G = V'DV,$$

where $V = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_m) \in \mathbb{R}^{n\times m}$, $D = (\lambda_1, \ldots, \lambda_n)$ and $\lambda_1 \geq \cdots \geq \lambda_n > 0$.
Let $X = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})V \in \mathbb{R}^{n\times m}$. The distances between $\mathbf{x}_1, \ldots, \mathbf{x}_m$ equal the prescribed dissimilarities. Indeed, since $\mathbf{x}_i = \sqrt{\lambda_i}\mathbf{v}_i$, we have

$$\begin{aligned} d(\mathbf{x}_i, \mathbf{x}_j)^2 &= (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j) \\ &= \mathbf{x}_i'\mathbf{x}_i + \mathbf{x}_j'\mathbf{x}_j - 2\mathbf{x}_i'\mathbf{x}_j \\ &= \lambda_i\mathbf{v}_i'\mathbf{v}_i + \lambda_j\mathbf{v}_j'\mathbf{v}_j - 2\lambda_i\lambda_j\mathbf{v}_i'\mathbf{v}_j \\ &= g_{ii} + g_{jj} - 2g_{ij} \\ &\quad \text{(by Lemma 4)} \\ &= a_{ii} + a_{jj} - 2a_{ij} = -2a_{ij} = \delta_{ij}^2. \end{aligned}$$

Since $G = H_m A H_m$ and H_m has an eigenvalue equal to 0 it is clear that G also has such an eigenvalue. Therefore, $rank(G) \leq m - 1$, so there exist m vectors of dimensionality not larger than $m - 1$ such that their distances are equal to the given dissimilarities.
We saw that the matrices XX' and $G = X'X$ have the same rank and their non-zero eigenvalues are positive numbers and have the same algebraic multiplicities for both matrices.

Let **w** be a principal component of the matrix $X' \in \mathbb{R}^{n \times m}$, that is, an eigenvector of the matrix XX'. Suppose that $rank(G) = r$ and let $X' = UDV'$ be the thin SVD decomposition of the matrix X', where $D = (\sigma_1, \ldots, \sigma_r)$ and $U, V \in \mathbb{R}^{m \times r}$. The matrices U and V have orthogonal columns, so $U'U = V'V = I_r$.
Since the numbers $\sigma_1, \ldots, \sigma_r$ are positive, D is invertible and we obtain $U = X'VD^{-1}$.
Thus, MDS involves a process that is dual to the usual PCA; some authors refer to it as the *dual PCA*.

MATLAB deals with metric MDS using the function `cmdscale`.

- The function call `X = cmdscale(D)` is applied to a distance matrix $D \in \mathbb{R}^{m \times n}$, and returns a matrix $X \in \mathbb{R}^{m \times n}$.
- The *rows* of X are the coordinates of m points in n-dimensional space for some n, where $n < m$. When D is a Euclidean distance matrix, the distances between those points are given by D.
- The number n is the smallest dimension of the subspace in which the m points whose inter-point distances are given by D can be embedded.

- `[X,e] = cmdscale(D)` also returns the eigenvalues of XX' as components of the vector e.
- When D is Euclidean, the first n elements of e are positive, the rest zero.
- If the first k elements of e are much larger than the remaining $n - k$, then it is possible to use the first k columns of X to produce k-dimensional vectors whose inter-point distances approximate D. This can provide a useful dimension reduction for visualization, e.g., for $k = 2$.

D need not be a Euclidean distance matrix.

- If it is non-Euclidean or a more general dissimilarity matrix, then some elements of e are negative, and `cmdscale` chooses n as the number of positive eigenvalues.
- The reduction to n or fewer dimensions provides a reasonable approximation to D only if the negative elements of e are small in magnitude.

Example

We start from a matrix of driving distances between five northeastern cities: Boston, Providence, Hartford, New York, and Concord.

$$\text{dist} = \begin{pmatrix} 0 & 41.9000 & 92.8800 & 189.9000 & 63.4700 \\ 41.9000 & 0 & 65.3600 & 154.8400 & 95.7800 \\ 92.8800 & 65.3600 & 0 & 99.7600 & 115.5900 \\ 189.9000 & 154.8400 & 99.7600 & 0 & 213.7800 \\ 63.4700 & 95.7800 & 115.5900 & 213.7800 & 0 \end{pmatrix}.$$

Example

Applying `cmdscale` to this matrix, `[X,e]=cmdscale(dist)` produces the matrix

$$X = \begin{pmatrix} 58.1439 & -20.4773 & -4.2664 \\ 19.3304 & -34.2586 & 3.4664 \\ -29.8485 & 8.8070 & 1.1787 \\ -129.6169 & 7.7975 & -1.1686 \\ 81.9911 & 38.1313 & 0.7899 \end{pmatrix}$$

and the matrix of eigenvalues

$$e = 1.0e+004 * \begin{pmatrix} 2.8168 \\ 0.3185 \\ 0.0034 \\ -0.0000 \\ -0.0006 \end{pmatrix}$$

Next, by defining the matrix

cities = ['BOS';'PRO';'HAR';'NYC';'CON']

the results are displayed using

plot(X(:,1),X(:,2),'+')
gname(cities)

which results in the representation shown next.

Representation of the five cities

The representation is approximative, but the relative positions of the cities is reasonably close to their real placement.